

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2005年4月28日 (28.04.2005)

PCT

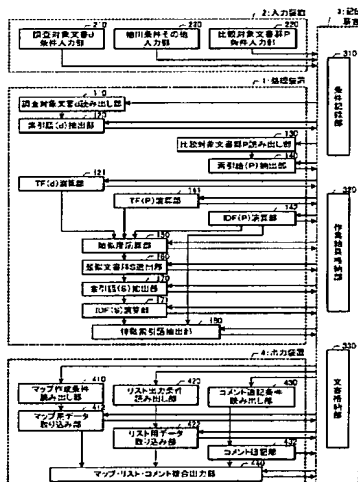
(10) 国際公開番号
WO 2005/038672 A1

- (51) 国際特許分類: G06F 17/30
- (21) 国際出願番号: PCT/JP2004/015082
- (22) 国際出願日: 2004年10月13日 (13.10.2004)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2003-393081
2003年10月21日 (21.10.2003) JP
特願 2004-160401 2004年5月31日 (31.05.2004) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社アイ・ピー・ビー (INTELLECTUAL PROPERTY BANK CORP.) [JP/JP]; 〒1050001 東京都港区虎ノ門 1-2 1-1 9 Tokyo (JP).
- (72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 増山 博昭 (MASUYAMA, Hiroaki) [JP/JP]; 〒5600054 大阪府豊中市桜の町 6-9-3 0-5 0 1 Osaka (JP). 佐藤 晴正 (SATO, Haru-tada) [JP/JP]; 〒1050001 東京都港区虎ノ門 1-2 1-1 9 株式会社アイ・ピー・ビー内 Tokyo (JP).
- (74) 代理人: 保坂 延寿 (HOSAKA, Nobuhisa); 〒3300074 埼玉県さいたま市浦和区北浦和 2-1 4-7 Saitama (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS,

/続葉有/

(54) Title: DOCUMENT CHARACTERISTIC ANALYSIS DEVICE FOR DOCUMENT TO BE SURVEYED

(54) 発明の名称: 調査対象文書の文書特徴分析装置



(57) Abstract: An index entry word extraction device includes: input means (1) for inputting a document-to-be-surveyed d and documents-to-be-compared P; index entry word extraction means (120) for extracting an index entry word from the document-to-be-surveyed d; first appearance frequency calculation means (142) for calculating a function value IDF (P) of the appearance frequency of the extracted index entry word in the documents-to-be-compared P; similar documents selecting means (160) for selecting similar documents S similar to the document-to-be-surveyed d in the documents-to-be-compared P according to the data on the document-to-be-surveyed d; second appearance frequency calculation means (171) for calculating the function value IDF (S) of the appearance frequency of the extracted index entry word in the similar documents S; and output means (4) for outputting each index entry and its positioning data according to the combination of the function values of the respective appearance frequencies in the documents-to-be-compared and the similar documents which have been calculated. Thus, it is possible to accurately grasp the feature of the document-to-be-surveyed.

(57) 要約: 調査対象文書 d 及び比較対象文書群 P を入力する入力手段 1 と、調査対象文書 d 内の索引語を抽出する索引語抽出手段 120 と、前記抽出された索引語の、比較対象文書群 P における出現頻度の関数値 IDF (P) を算出する第 1 出現頻度算出手段 142 と、調査対象文書 d のデータに基づき、比較対象文書群 P の中から調査対象文書 d に類似する類似文書群 S を選出する類似文書群選出手段 160 と、前記抽出された索引語の、類似文書群 S における出現頻度の関数値 IDF (S) を算出する第 2 出現頻度算出手段 171 と、各索引語についての、前記算出された比較対象文書群及び類似文書群における各出現頻度の関数値の組合せに基づき、各索引語とその位置づけデータとを出力する出力手段 4 と、を備える。これにより、調査対象文書の性格的的確な把握を可能にする。

WO 2005/038672 A1



LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA,
NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE,
SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, YU, ZA, ZM, ZW.

IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF,
BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN,
TD, TG).

- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE,

添付公開書類:

— 国際調査報告書

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。